Explanation Production by Expert Planners

Susan Bridges
James D. Johannes

University of Alabama in Huntsville
Computer Science Department
Huntsville, AL 35899

Abstract

Although the explanation capability of expert systems is usually listed as one of the distinguishing characteristics of these systems, the explanation facilities of most existing systems are quite primitive. Computer generated explanations are typically produced from canned text or by direct translation of the knowledge structures. Explanations produced in this manner bear little resemblance to those produced by humans for similar tasks.

The focus of our research in explanation is the production of justifications for decisions by expert planning systems. An analysis of justifications written by people for planning tasks has been taken as the starting point for our research. The purpose of this analysis is two-fold. First, analysis of the information content of the justifications will provide a basis for deciding what knowledge must be represented if human-like justifications are to be produced. Second, an analysis of the textual organization of the justifications will be used in the development of a mechanism for selecting and organizing the knowledge to be included in a computer-produced explanation.

This paper describes a preliminary analysis that has been done of justifications written by people for a planning task. It is clear from this analysis that these justifications differ significantly from those that would be produced by an expert system by tracing the firing of production rules. The results from the text analysis have been used to develop an augmented phrase structured grammar (APSG) that describes the organization of the justifications. The grammar was designed to provide a computationally feasible method for determining textual organization that will allow the necessary information to be communicated in a cohesive manner.

Introduction

Expert system technology has made impressive strides in recent years. Simple rule-based architectures have given way to sophisticated hybrid systems that support a variety of knowledge representation and reasoning mechanisms. One aspect of "expert performance" that has lagged in development is that of explanation. Although most expert system development tools include facilities for building elaborate graphic interfaces, there are many explanation tasks that are not amenable to graphic presentation. The need for textual explanation seems evident when one considers the vast amount of written documentation that human experts are expected to provide to justify the decisions they make.

The expert system literature generally cites three main purposes for explanation facilities [1, 4, 6, 13]. First, explanation can be used by the knowledge engineer to test and debug the system. Second, explanation assures sophisticated users that the system's knowledge and reasoning process are sound and allows the detection of situations in which the system is being asked to perform a task outside the boundaries of its capability. Finally, explanation facilities can be employed to instruct naive users about the knowledge of the system. These functions relate to the interaction of the user and expert system in the course of a consultation. Another aspect of explanation that will become more important as expert systems are used in complex and critical domains is that of recording justifications of decisions. Human experts are usually called on to provide a written justification for the validity of their decisions. It would appear reasonable, therefore, that a computer program that aids in making these decisions should also provide assistance in providing justifications for the decisions.

Expert system explanations have typically been produced by the use of canned text (for tasks such as defining terms) or by tracing the rules that have fired during the inference process. Systems that generate rule traces usually have some provision for translating the syntax of the rules into a natural language form for presentation to the user. Although this approach offers more explanation capability than is found in traditional computer programs, the traces produced are very different from the explanations one would expect from a human expert. In particular, rule traces are not adequate for answering questions of the form "Why is this a valid decision?" The research described in this paper is based on the observation that when human experts are asked to justify decisions they

have made, they do not merely recite the steps taken in reaching the decision.

The limitations of the "trace the rules" approach have been widely discussed in the literature [3, 12, 14]. These limitations are of three types. In the first kind, the traces tend to be very long and contain much information that is of no interest to the user. This problem becomes more acute as the size of the system grows. The second problem with rule traces is the absence of much information that would be expected in a human-produced explanation. For example, information about the problem solving strategy of the system is not present in the rule trace because it is not explicitly represented in the rules and knowledge structures of the system. Other information, such as that needed to define terms, is often absent from the knowledge base altogether. The third type of problem concerns the structure of the explanations produced. Research in natural language processing has shown that multisentential text produced by humans exhibits a characteristic structure and organization that facilitates the communication process [7, 8]. This structure is not found in rule traces.

Much of the research in explanation production done to date has been in systems that perform medical diagnosis or fault diagnosis in electronic or mechanical devices. This research has addressed the first two problems listed above by developing methods for incorporating the knowledge needed for explanation in the knowledge structures of the system and for tailoring the system's responses to the user [3, 13]. The problem of structuring explanatory text has been largely unaddressed.

The goal of our current research is the development of a methodology for selecting and organizing the knowledge that is to be used to construct multisentential explanatory text that is a justification of the recommendations made by an expert planning system. The task of natural language generation has traditionally been divided into two components. A strategic component determines the content and structure of the text while a tactical component determines the natural language surface structure (which words and syntactic structures to use). The emphasis of our research is the strategic component.

Expert planning programs offer an attractive test bed for the production of justifications of decisions. There has been a great deal of research into the development of general frameworks for planning systems [2, 9], but this work has largely ignored questions of explanation.

Analysis of Text

Students are taught from early elementary years that there are ways to organize writing that will increase its effectiveness. It stands to reason, then, that text written for a specific purpose in a limited domain will exhibit more regularity of organization than text in general. Language understanding systems have long made use of this property [15]. For this reason, it was thought that the analysis of text written for a specific type of task could be used as the basis for the development of a grammar formalism that could be used in the generation of similar text by an expert system.

Written, rather than spoken text, was used in the analysis phase of this research for several reasons. First, written text is generally better planned and organized than spoken text. Spoken text often contains partial sentences and ungrammatical constructions that would be unacceptable in written form. In addition, speakers use facial expressions and tone of voice to convey much of their meaning. Spoken text is often directed toward a more specific audience and so requires a more elaborate user model than written text.

The type of task chosen is that of justifying the validity of a plan constructed by an expert planning system. An initial text analysis was done using justifications of Master's degree plans of study written by University of Alabama in Huntsville graduate students. Additional analysis has since been done of justifications of travel itineraries. The justifications of plans seem to follow the general form of identifying the each component of the plan, and the planning constraints that each component satisfies.

A Grammar Describing the Structure of Justifying Text

The augmented phrase structured grammar (ASPG) formalism was chosen as a representation for the text structure. Other approaches to representing text structure that have been used in previous research include fixed semantic patterns [3, 10, 11] and context-free grammars [8, 14]. The use of semantic patterns limits the flexibility of the representation by restricting the number of text structures that can be generated to a small finite set. Although the use of a context-free grammar allows an infinite number of structures to be generated, a

mechanism outside the grammar itself must be used to control the application of rewriting rules when more than one applies. The APSG facilities for attaching attributes to non-terminals and conditions to rewriting rules allows generation of an infinite number of structures and provides a method for embedding the control of the application of rewriting rules in the grammar.

After study of the justifications submitted, an attempt was made to develop an APSG that could be used to guide the generation of similar justifications by an expert system. The starting point for text generation is necessarily some representation of the relevant knowledge. It is understood that an explanation system cannot communicate information that the knowledge base does not contain or cannot derive. Thus, one important factor in the development of these systems is the representation of appropriate knowledge. It is assumed in this paper, that the appropriate knowledge is represented, but the method of representation is left undefined.

The APSG formalism uses auxiliary evaluation functions in the assignment of values to attributes and testing of conditions on rewriting rules[15]. In order to use the knowledge of the system to direct the generation of explanations, it was necessary to provide an interface between the grammar and the knowledge base. This is done by use of special auxiliary evaluation functions that access the knowledge base and return values that can be used to direct the application of grammar rules and as building blocks in the message constructed by the generation process. In addition to inherited and synthesized attributes, attributes that derive their values solely from functions that access the knowledge base are called assigned attributes. The Start symbol of the grammar will have one or more attributes that are given values before the generation process begins.

It is assumed that the generation process will proceed in a left to right, depth-first manner and so the restriction is imposed that synthesized attributes can only be inherited from left to right. In addition, the generation process will initially proceed in a top down manner and so attributes cannot be synthesized in a sub-tree and then used in a condition at the root of the sub-tree to test its validity. This means that attributes used in conditions will, in general, be inherited or assigned. When more than two rules can be applied, it is assumed that the conditions on the rules are sufficient to decide which is applicable so that backtracking can

be avoided. Evaluation rules for attributes that will be tested in conditions are given before the rewriting rules in which the conditions are tested.

The grammar given is unique in that it does not have any terminal symbols. It is assumed that "non-terminal" symbols, once generated, are never retracted. Thus, when a non-terminal is encountered for which there is no applicable rule, that symbol, in effect, becomes a terminal symbol. The tree built as the rules are applied acts as a framework for the message to be built. Pieces of the message are built at each of the leaves and are brought together as as synthesized attributes from the leaves to the root. The message this built has a structure imposed by embedded lists and can contain additional information such as focus of attention, tense, etc.

No attempt has been made to construct a grammar that can generate all of the organizations found in the justifications that were studied. Rather, the grammar is an attempt to provide a method for providing an organization that is computationally feasible and yet flexible enough to allow the necessary information to be communicated in a planned cohesive manner.

Summary

Human experts are often expected to provide written justification of the decisions they make. As the use of expert systems becomes more widespread, it will be become increasingly important for these systems to have the capability to compose text that justifies their decisions. Traditional "trace of the rules" explanations are not sufficient for this task because they include much information that not pertinent, they omit other information that is typically found in human-provided explanations, and they lack any organizing structure. This paper explores the possibility of using an augmented phrase structured grammar to describe the structure of justifications of expert planning decisions. The grammar provides a mechanism selecting and organizing the information to be provided in the justification.

References

1. Buchanan, Bruce G., and Shortliffe, Edward H. *Rule-Based Expert Systems*. Addison-Wesley, Reading, MA, 1984.

2. Chapman, David. Planning for Conjunctive Goals, *Artificial Intelligence*, 32, 1987, 333-377.

3. Clancey, William J. The Epistemology of a Rule-Based Expert System--a Framework for Explanation. *Artificial Intelligence*, 20, 1983, 215-251.

4. Finin, Timothy W., Joshi, Aravind K., and Webber, Bonnie Lynn. Natural Language Interactions with Artificial Experts. *Proceedings of the IEEE*, 74, 1986, 921-938.

5. Hasling, D. W., Clancey, W. J., and Rennels, G. Strategic Explanations for a Diagnostic Consultation System. *International Journal of Man-Machine Studies*, 20, 1984, 3-19.

6. Hayes-Roth, Frederick, Donald A. Waterman, and Douglas B. Lenat, (Eds), *Building Expert Systems*. Addison-Wesley Publishing Company, Reading, MA, 1983.

7. Mann, William C., and Thompson, Sandra A. Rhetorical Structure Theory: Description and Construction of Text Structures. In: Gerard Kempen (Ed.), *Natural Language Generation*, Martinus Nijhoff Publishers, Dordrecht, 1987, 85-96.

8. McKeown, Kathleen R. *Text Generation*, Cambridge University Press, Cambridge, UK, 1985.

9. Sacerdoti, Earl D. *A Structure for Plans and Behavior*, SRI Technical Report TN-109 SRI International, Menlo Park, CA, 1975.

10. Schank, Robert C., *Explanation patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

11. Schulman, Robert, and Hayes-Roth, Barbara, Plan-Based Construction of Strategic Explanations. Knowledge Systems Laboratory, Report No. KSL 88-23, Stanford University, Stanford, CA, 1988.

12. Swartout, William R., XPLAIN: a System for Creating and Explaining Expert Consulting Systems, *Artificial Intelligence*, 21, 1983, 285-325.

13. Swartout, William R. Knowledge Needed for Expert System Explanation, *Future Computing Systems*, 1, 1986, 99-114.

14. Weiner, J. L. BLAH, A System Which Explains its Reasoning. *Artificial Intelligence*, 15, 1980, 19-48.

15. Winograd, Terry, (1983) *Language as a Cognitive Process: Syntax*, Addison-Wesley, Reading, MA.